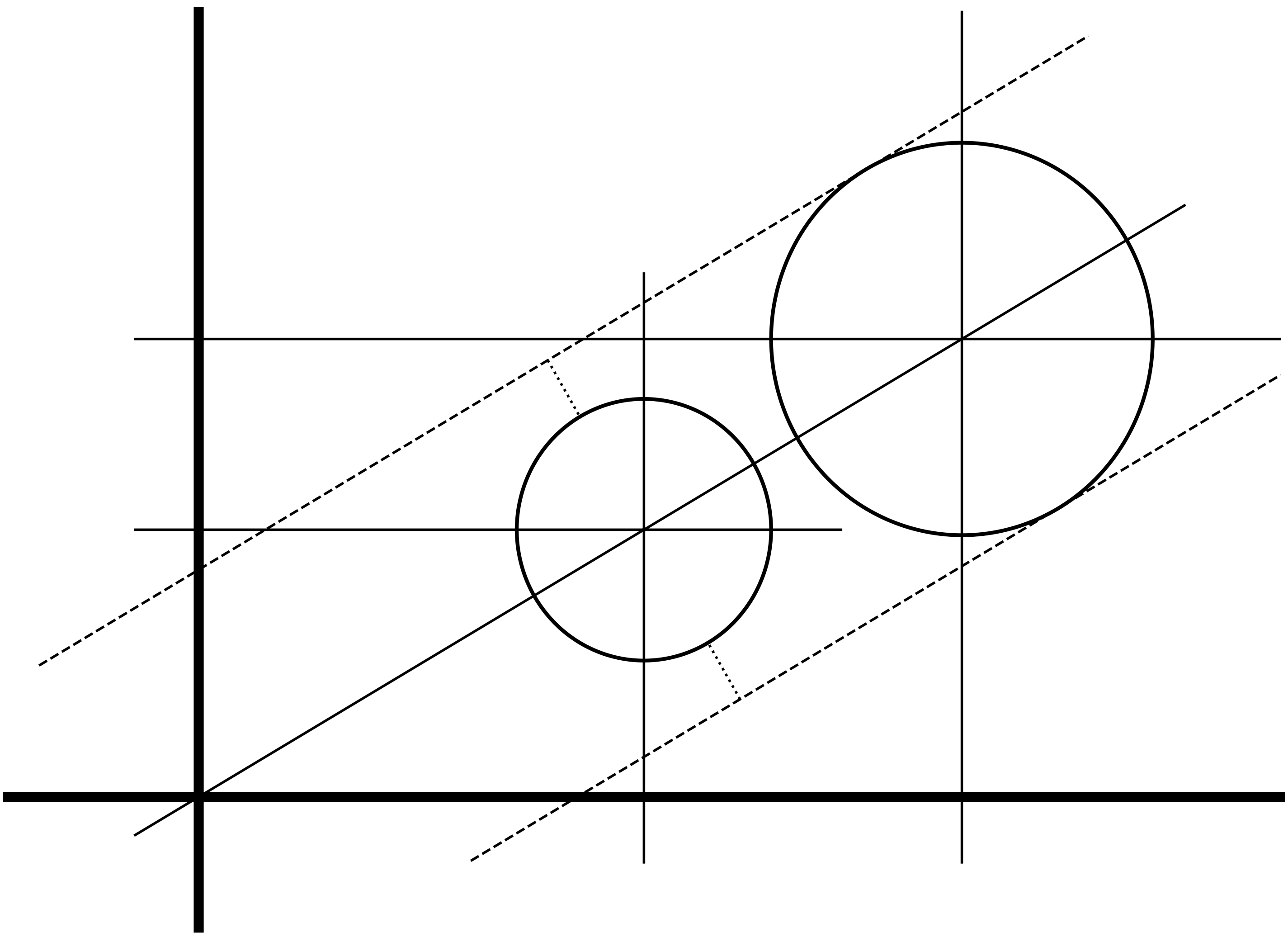
Confidence Sets based on Shrinkage

A.C. Bor



Confidence Sets based on Shrinkage

by

A.C. Bor

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Friday June 27th, 2025 at 10:00 AM.

Student number:	5616565	
Project duration:	April 1, 2025 – June 27, 2025	
Thesis committee:	D r. J. Söhl,	TU Delft, supervisor
	Dr. O. Jaibi,	TU Delft

An electronic version of this bachelor thesis is available at `http://repository.tudelft.nl/`.

Contents

Layman's Summary

Many real-world and theoretical phenomena can be modelled by a multivariate normal distribution, so understanding how to make reliable statements about its unknown mean vector $\theta \in \mathbb{R}^k$ is a central problem in statistics. Traditionally we tackle this in three stages:

1. **Estimation**: choose a rule (an estimator) that produces an educated guess for θ based on our data.
2. **Hypothesis testing**: decide whether the data provides enough evidence to reject a given hypothesis about θ.
3. **Confidence sets**: construct a region in $\mathbb{R}^k$ that, with probability $1-\alpha$, contains the true θ.

A surprising result by Charles Stein shows that when $k \geq 3$, the usual sample mean is *inadmissible* under mean-squared error: by shrinking the raw data vector toward a fixed point (typically the origin), one can obtain strictly smaller average squared error (Stein, 1954). James and Stein made this concrete with the *(positive-part) James–Stein estimator* (Baranchik, 1964, James and Stein, 1961), which multiplies the data vector X by

$$\left(1 - \frac{a}{||X||^2}\right)_+, \qquad a > 0.$$

Although this introduces some bias, it cuts variance enough to improve overall accuracy whenever $k \geq 3$.

This thesis explores using that same shrinkage idea not just for point estimation, but for building better hypothesis tests and confidence sets:

- **Chapter 1**: Derive the sampling distribution of $T(X) = \|(1 - a/\|X\|^2)_+ X - \theta\|^2$. We give closed-form expressions for any spherically symmetric distribution centred around θ.
- **Chapter 2**: Use this distribution to perform
  - **Simple tests**: testing $\theta = \theta_0$,
  - **Composite tests**: testing whether θ lies in a region,

  at a chosen significance level α.
- **Chapter 3**: Construct $(1-\alpha)$-confidence sets for θ centred at the positive-part James–Stein estimator by
  1. the *plug-in method* (estimating $\|\theta\|$ via various rules and solving for the radius), and
  2. *test-inversion* (collecting all y not rejected by the James–Stein test of "$\theta = y$" at level α).

Numerical approximations show that the quantiles of the positive part James-Stein test statistic are smaller than that of the usual chi-squared test statistic.

Abstract

We study the problem of constructing confidence sets for the mean vector θ of a k-variate spherically symmetric distribution by centring them at the positive-part James–Stein estimator. Exploiting its superior risk properties whenever $k \geq 3$, we first derive in Chapter 2 the exact sampling law of the test statistic

$$T(X) = \|\hat{\theta}^{+}_{JS} - \theta\|^2,$$

showing it consists of a point mass at $\|\theta\|^2$ and a continuous density component valid for any spherically symmetric model. In Chapter 3, we develop level-α procedures for both simple and composite hypotheses about θ, illustrated by a worked example with $k = 4$, $\alpha = 0.05$. Chapter 4 then inverts these tests to form $(1 - \alpha)$-confidence sets via two approaches: (i) a plug-in method using various norm estimators (including the James–Stein shrinkage itself) and (ii) a test-inversion principle guaranteeing exact coverage. Numerical comparisons confirm that the plug-in method produces smaller radii than classical sample-mean–centred sets. Our work thus extends classical multivariate inference by integrating shrinkage estimation into confidence-set theory for spherical distributions.

1
Introduction

It is well known that many processes tend to follow a normal distribution, thanks in large part to the Central Limit Theorem, and thus properties of this distribution are of high importance within mathematical statistics. One such property we are often times interested in estimating is its mean $\theta \in \mathbb{R}^k$. Since the normal distribution is a symmetric distribution it would be reasonable to assume that the sample mean would be the best estimator for θ but Charles Stein showed in 1954 that this estimator is actually inadmissible in 3 or more dimensions (Stein, 1954). In a later paper in collaboration with Willard James a concrete estimator with a lower mean squared error was introduced which dubs the name the James-Stein estimator (James and Stein, 1961). In the same paper a slight modification was conjectured to perform even better and this was shows to be true by Baranchik, this estimator is known as the positive part James-Stein estimator (Baranchik, 1964).

Whilst a lot of excellent research has been done with respect towards the positive part James-Stein estimator an area that remains to be explored more are confidence sets centred around this estimator. Samworth showed some promising work regarding such confidence sets using the bootstrap method and also provided some analytical results which involved finding the first few terms of the Taylor series of the law of the test statistic centred around the origin (Samworth, 2005).

In this work we are mainly concerned with analytical results. Chapter 2 starts by formally defining the problem at hand followed by some properties of the defined quantities. Next the chapter focusses on the problem of finding a general density function for the norm of X for arbitrary spherically distributed random variables centred at θ. The chapter closes with its main result, the law of the positive part James-Stein test statistic, and some of its properties. At the end some validation is performed by taking the limit of a t-distribution and showing that this converges to a normal distribution as expected.

Chapter 3 elaborates on classical hypothesis testing. It starts with an exploration of some of the properties of the previously investigated positive part James-Stein test statistic with regards to hypothesis testing. This is followed by a small discussion on simple and general hypothesis testing. The chapter closes with a worked example.

The final chapter, Chapter 4, applies previously found results to construct confidence sets for θ. It starts of by showing how classic confidence sets are constructed after which we will try to generalise this to our estimator. We do this by estimation of the norm of the parameter θ for which we have selected several candidates. The chapter also discusses the process of constructing confidence sets by, in some sense, inverting the hypothesis tests.

2

The Law of the positive part James-Stein Test Statistic

In this section we will be exploring the distribution of the distance between the true mean of some spherically distributed probability distribution and the positive part James-Stein estimator. We first formalise these concepts and from there build towards the main result of this section.

2.1. Notation and Core Concepts

As stated before we focus on spherically distributed random variables of which we now give a formal definition.

Definition 2.1.1. A k-dimensional random variable X with mean $\theta \in \mathbb{R}^k$ is said to be *spherically symmetric around* θ, spherically distributed for short, if for all orthogonal matrices Q we have

$$\big(X-\theta\big) \sim Q\big(X-\theta\big),$$

where $\sim$ denotes equal in distribution.

Intuitively this can be interpreted as that the probability of X taking some value in $\mathbb{R}^k$ only depends on the distance from its true mean θ. Mathematically this means that we can write $f_X(x) = g(||x-\theta||^2)$ for some function $g : [0,\infty) \to [0,\infty)$ where f_X denotes the probability density of X. We take $k \geq 3$ and note it is tempting to say that whenever a random variable is spherically distributed, then its entries must be independent. This is not only false but actually characterises the k-variate normal distribution.

Theorem 2.1.2. *Let X be a k-dimensional spherically distributed random variable with mean θ and mutually independent components $X_1,...X_k$. Then $X \sim N_k(\theta, \sigma^2 I_k)$ for some $\sigma^2 \geq 0$.*

Proof. Define $Y := X-\theta$. Let $\phi(t) = \prod_{j=1}^k \phi_j(t_j)$ be its characteristic function and note that for all orthogonal matrices Q we have $\phi(t) = \phi(Qt)$. Since ϕ only depends on the norm of its argument we can write $\phi(t) = \psi(||t||)$ for some function ψ. If we set $t_3 = ... = t_k = 0$ then we get

$$\psi\Big(\sqrt{t_1^2+t_2^2}\Big) = \phi_1(t_1)\phi_1(t_2)\prod_{j=3}^{k}\phi_1(0) = \phi_1(t_1)\phi_1(t_2),$$

in which we used that all marginal characteristic functions are equal. Take the natural log on both sides, define $f(s) := \ln\phi_1(s)$ and $u(s) := f(\sqrt{s})$. Then

$$f(x)+f(y) = f\Big(\sqrt{x^2+y^2}\Big) \implies u(r)+u(s) = u(r+s).$$

The functional equation for u is known as Cauchy's equation and has unique solution of the form $u(r) = cr$ for some real constant c. This implies that we have $\phi_1(t) = \exp\{f(t)\} = \exp\{u(t^2)\} = \exp\{ct^2\}$. Finally since $|\phi_1(t)| \leq 1$ we have $c \leq 0$ so we can write $\phi_1(t) = \exp\{-\frac{1}{2}\sigma^2 t^2\}$ for some $\sigma^2 \geq 0$ which uniquely defines a normal distribution.

□

Note that Theorem 2.1.2 also allows the variance of our normal distribution to be equal to zero. In this case we say that X is degenerate and thus by definition assumes one value with probability one.

Charles Stein showed that for k-variate normal distributions the sample mean is inadmissible (Stein, 1954). An estimator with a lower mean squared error was found in a later paper in collaboration with Willard James together with an estimator which was conjectured to perform even better. This was shown to be true by Baranchik (Baranchik, 1964).

Definition 2.1.3. (Baranchik, 1964, James and Stein, 1961) Let X be a k-variate spherically distributed random variable with true mean θ. *The positive part Stein estimator* for θ, denoted by $\hat{\theta}^+_{JS}$ is defined by

$$\hat{\theta}^+_{JS} = \left(1 - \frac{a}{||X||^2}\right)_+ X,$$

in which $(\cdot)_+ = \max\{0, \cdot\}$, $||X||$ the usual euclidean norm and a the *shrinkage factor*.

Since the positive part James-Stein estimator performs better than the sample mean under the mean squared error it makes sense to use it for other purposes than estimation. Later chapters will elaborate on this but for now, we need one more definition.

Definition 2.1.4. Let X be a k-variate spherically distributed random variable around θ. *The positive part James-Stein test statistic* is given by

$$T(X) = \left|\left|\left(1 - \frac{a}{||X||^2}\right)_+ X - \theta\right|\right|^2.$$

Note that this is a one-dimensional random variable.

Considering the form of the positive part James-Stein test statistic it seems reasonable to assume that its probability distribution only depends on θ through its norm, $||\theta||$. It turns out that this is a general result for test statistics centred around estimators of a certain form.

Proposition 2.1.5. *(Samworth, 2005) For $\alpha \in (0,1)$, the upper α-point of the sampling distribution of $||\gamma(||X||)X - \theta||^2$ depends on θ only through its norm $||\theta||$.*

A consequence of this is that regardless of how the length of θ is distributed among its entries the distribution of the positive part James-Stein test statistic will behave the same way. It also gives us a quick way to check whether or not our found distribution functions are plausible or not.

2.2. Distribution of $||X||$

In the previous section we stated that if X is spherically distributed around θ, then there must exist some function g_X such that we have $f_X(x) = g_X(||x - \theta||^2)$ but one might ask how we would find such a function. This question is equivalent to asking what is the density of the radius of X from the origin. In some cases these functions can easily be derived from their k-variate counterparts. When we have $X \sim N_k(\theta, I_k)$ this function is given by

$$f_X(x) \propto \exp\Big\{-\frac{1}{2}\big(x-\theta\big)^T I_k\big(x-\theta\big)\Big\} \implies g_X(u) \propto \exp\Big\{-\frac{1}{2}u\Big\},$$

The problem of finding such a function g_X becomes quite a bit more difficult when we drop the centrality condition. For the norm of X as defined above its distribution is actually known as the non-central chi distribution, beware that this is the root of the non-central chi-squared distribution. Its density is given by

$$f_{||X||}(x) = \frac{e^{-(x^2+\lambda^2)/2} x^k \lambda}{(\lambda x)^{k/2}} I_{k/2-1}(\lambda x), \tag{2.2.1}$$

in which we used $\lambda = ||\theta||$ and $I_\nu(z)$ is the modified Bessel function of the first kind.

Though the normal distribution is perhaps the most common it would be preferable if we had a general expression for the distribution of $R = ||X||$ for arbitrary spherically distributed random variables. It turns out there exists such a form.

Lemma 2.2.1. *Let X be a k-variate spherically distributed random variable around θ with density function $f_X(x) = g_X(||x-\theta||^2)$ and let λ denote the norm of θ. The density of $R = ||X||$ is given by*

$$f_R(r) = r^{k-1}\Big|S^{k-2}\Big|\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}} g\big(r^2+\lambda^2-2r\lambda t\big)dt.$$

Where $|S^{k-2}|$ denotes the surface area of a $(k-2)$-dimensional sphere.

Proof. To find the probability that $R = r$ we integrate over the sphere with radius r. This gives

$$f_R(r) = \int_{||x||=r} f_X(x)dS(x) = \int_{||x||=r} g\big(||x-\theta||^2\big)dS(x),$$

where $dS(x)$ denotes the $k-1)$-dimensional surface element. We substitute polar coordinates along θ which gives us $x\cdot\theta = r||\theta||\cos(\varphi)$ and $||x-\theta||^2 = r^2 + ||\theta||^2 - 2r||\theta||\cos(\varphi)$. Write $x = ru, u \in S^{k-1} \subset \mathbb{R}^k$ then we have

$$dS(x) = r^{k-1}dS(u) = r^{k-1}\Big|S^{k-2}\Big|\big(\sin(\varphi)\big)^{k-2}d\varphi$$

hence we can write

$$f_R(r) = r^{k-1}\Big|S^{k-2}\Big|\int_0^{\pi} g\big(r^2+||\theta||^2-2r||\theta||\cos(\varphi)\big)\big(\sin(\varphi)\big)^{k-2}d\varphi.$$

Finally let $t = \cos(\varphi)$ and we have

$$f_R(r) = r^{k-1}\Big|S^{k-2}\Big|\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}} g\big(r^2+||\theta||^2-2r||\theta||t\big)dt$$

as desired.

□

Even though this integral cannot always be solved analytically it does provide us with a closed form expression for the density of R provided we know g_X. This integral may be approximated using a number of numerical methods which are beyond the scope of this work.

Lemma 2.2.1 does allow us to verify some previous results, let $X \sim N_k(\theta, I_k)$ then we have

$$f_R(r) = \frac{r^{k-1}}{(2\pi)^{k/2}} \left|S^{k-2}\right| \int_{-1}^{1} (1-t^2)^{\frac{k-3}{2}} \exp\left\{-\frac{1}{2}\left(r^2+\lambda^2-2r\lambda t\right)\right\} dt.$$

Simplifying by taking out all the non t-dependent parts and substituting the surface area of a $(k-2)$-dimensional unit sphere we get

$$f_R(r) = \frac{e^{-(x^2+\lambda^2)/2} r^{k-1}}{(2\pi)^{k/2}} \left(\frac{2\pi^{k/2-1}}{\Gamma((k-1)/2)}\right) \int_{-1}^{1} (1-t^2)^{\frac{k-3}{2}} e^{r\lambda t} dt.$$

The integral in this expression is a result known as the Beta integral and evaluates to

$$\int_{-1}^{1} (1-t^2)^{\frac{k-3}{2}} e^{r\lambda t} dt = \frac{\sqrt{\pi}\Gamma((k-1)/2)}{(\frac{1}{2}\lambda r)^{k/2-1}} I_{k/2-1}(\lambda r),$$

of which the derivation can be found in Appendix A. When we combine all this we get exactly Equation 2.2.1 as desired.

Before we close this subsection we introduce one final seemingly unrelated result. Provided we know the distance of X from the origin we still need more information to uniquely determine the actual point X in space where we by uniquely mean unique in probability, i.e. two points A and B are equivalent if they have equal probability. There are of course several ways to do this but we will be using the angle between θ and X.

Lemma 2.2.2. *Let X be a k-variate spherically symmetric distributed random variable with radial density function $g(||x-\theta||^2)$, define V as $\langle X/||X||, \theta/\lambda\rangle$ and take R to be the radius of X. Then the joint density of V and R is given by*

$$f_{R,V}(r,v) = r^{k-1}(1-v^2)^{\frac{k-3}{2}}|S^{k-2}|g(r^2+\lambda^2-2r\lambda v).$$

Note that V is the cosine of the angle between θ and X with respect to the origin.

Proof. Since our random variable is spherically distributed we may assume that θ is aligned with the first coordinate axis. We write all $x \in \mathbb{R}^k$ in polar form $x = ru$ where $r = ||x|| > 0$ and $u \in S^{k-1}$. Provided we know its spherical density function we may write

$$f_X(x) = g_X(||x-\theta||^2) = g_X(r^2+\lambda^2-2\lambda r u_1).$$

In polar coordinates we have $dx = r^{k-1} dr d\sigma(u)$ where $\sigma(u)$ is the surface-area measure on S^{k-1}. The joint density of $U = X/||X||$ and $R = ||X||$ is then

$$f_{U,R}(u,r) = r^{k-1} g(r^2+\lambda^2-2\lambda r u_1).$$

We now introduce

$$V = \left\langle \frac{X}{||X||}, \frac{\theta}{||\theta||}\right\rangle = \langle U, e_1\rangle = u_1 \in [-1,1].$$

For a fixed v the set $\{u \in S^{k-2} : u_1 = v\}$ is $(k-2)$-dimensional sphere with radius $\sqrt{1-v^2}$ with surface area

$$\left|S^{k-2}\right|(1-v^2)^{\frac{k-3}{2}}.$$

Combining all gives the density.

□

Since by definition if R and V have joint density $f_{R,V}$ the densities of R and V may be found by integrating out the other. Since V is the cosine of some real angle we have that $V \in [-1,1]$ and by integrating over that interval we get back Lemma 2.2.1. Though not necessarily useful we also state the following Corollary without proof.

Corollary 2.2.3. *Let X be a k-variate spherically symmetric distributed random variable with mean θ and density function $g(||x-\theta||^2)$. Let $Y = X/||X||$ and $\lambda = ||\theta||$. Define $V = \langle Y, \theta/\lambda \rangle$, then V has the following density*

$$f_V(v) = (1-v^2)^{\frac{k-3}{2}} \left|S^{k-2}\right| \int_0^\infty \rho^{k-1} g(\rho^2 + \lambda^2 - 2\lambda\rho v)\, d\rho.$$

Where $|S^{k-2}|$ denotes the surface area of a $(k-2)$-dimensional unit-sphere.

2.3. The Law of the positive part James-Stein Test Statistic

In Definition 2.1.4 we introduced the shrinkage factor a. In many works this factor is taken to be equal to $k-2$ or is adjusted depending on the known (or estimated) variance of the distribution at hand. We will assume arbitrary strictly positive a. For convenience we restate the positive part James-Stein test statistic

$$T(X) = ||\gamma(||X||)X - \theta||^2, \qquad \gamma(||X||) = \Big(1 - \frac{a}{||X||^2}\Big)_+. \tag{2.3.1}$$

The first observation we can make is that $\gamma(||X||) = 0$ when we have that $1 - a/||X||^2 \leq 0$ which in turn implies $||X|| \leq \sqrt{a}$. This is exactly our first result.

Proposition 2.3.1. *The positive part James-Stein test statistic is equal to $||\theta||^2$ if and only if $||X|| \leq \sqrt{a}$, i.e.*

$$\mathbb{P}\Big\{T(X) = ||\theta||^2\Big\} = \mathbb{P}\Big\{||X|| \leq \sqrt{a}\Big\} = \int_0^{\sqrt{a}} f_R(s)ds.$$

In which f_R is the density of the radius $||X||$.

If X follows a central k-variate normal distribution we have that by definition $||X||^2 \sim \chi_k^2(\lambda^2)$ and then this integral simplifies to the a-quantile of the non-central chi-squared distribution. Similar statements can be made for several known distributions, one of which was covered in the previous section. Another observation that may be made from Proposition 2.3.1 is that since the probability that $T(X)$ assumes the singular value $||\theta||^2$ is not equal to zero we are dealing with a mixed density with continuous part and a point mass. We are now ready to state the main result of this section.

Theorem 2.3.2. *Let X be a k-variate spherically distributed random variable around θ. The positive part James-Stein test statistic as defined in Definition 2.1.4 has probability density function*

$$f_T(t) = \Big(\int_0^{\sqrt{a}} f_R(s)ds\Big)\delta(t-\lambda^2) + \frac{1}{2\lambda}\int_{r_-(t;\lambda)}^{r_+(t;\lambda)} \frac{r}{r^2-a} f_{V,R}(v(r,t),r)dr,$$

in which

$$v(r,t) = \frac{(r^2-a)^2 - (t-\lambda^2)r^2}{2r(r^2-a)\lambda},$$

$f_{V,R}$ as in Lemma 2.2.2, δ the Dirac-Delta function and the lower and upper bounds are given by

$$r_\pm(t;\lambda) = \frac{1}{2}\Big(|\lambda \pm \sqrt{t}| + \sqrt{|\lambda \pm \sqrt{t}|^2 + 4a}\Big),$$

respectively.

The proof can be found in Appendix B. This is a very powerful result because it allows us to calculate actual quantiles of our test statistic $T(X)$ provided we know λ. Similar to before if we restrict ourselves to the normal distribution we can perform some simplifications. When $X \sim N_k(\theta, I_k)$ the distribution of V is actually known as the von Mises-Fisher distribution and we get the following form.

Corollary 2.3.3. *Let $X \sim N_k(\theta, I_k)$. Then the positive part James-Stein test statistic has the following density*

$$f_T(t) = \mathbb{P}\big(R \leq \sqrt{a}\big)\delta\big(t-\lambda^2\big) + C_k\int_{r_-(t;\lambda)}^{r_+(t;\lambda)} \frac{r^k}{r^2-a}\exp\Big\{\frac{r^2(\lambda^2-t)}{2(r^2-a)}\Big\}\big(1-v^2(r,t)\big)^{\frac{k-3}{2}} dr,$$

with

$$C_k = \frac{e^{-(\lambda^2+a)/2}}{\sqrt{\pi}\Gamma\big(\frac{1}{2}(k-1)\big)\lambda 2^{k/2}}, \qquad v(r,t) = \frac{(r^2-a)^2-(t-\lambda^2)r^2}{2r(r^2-a)\lambda},$$

in which r_-, r_+ are as defined in Theorem 2.3.2 and R follows a non-central chi-squared distribution. When $\lambda = 0$ then the density is given by

$$f_T(t) = \delta(t)F_{\chi^2_k}(a) + \frac{2^{-k/2}}{\Gamma(k/2)}\frac{r^k(t)}{\sqrt{t}\sqrt{t+4a}}\exp\Big\{-\frac{r^2(t)}{2}\Big\},$$

with

$$r(t) = \frac{1}{2}\Big(\sqrt{t}+\sqrt{t+4a}\Big).$$

Corollary 2.3.3 follows from a straight substitution in Theorem 2.3.2. The zero-mean case can be found using a limit, or more easily by modifying the proof of Theorem 2.3.2 as below.

Proof. Let $X \sim N_k(0, I_k)$. Then we can write

$$T(X) = \begin{cases} 0, & \text{if } ||X||^2 \le a, \\ \frac{(||X||^2-a)^2}{||X||^2}, & \text{if } ||X||^2 > a. \end{cases}$$

Which then implies we have

$$\mathbb{P}(T(X)=0) = \mathbb{P}(||X||^2 \le a) = F_{\chi^2_k}(a).$$

Let $u = ||X||^2$ and for values of $t > 0$ introduce the function

$$g(u) = \frac{(u-a)^2}{u} = t \implies u^2 - (t+2a)u + a^2 = 0.$$

In the region $u > a$ and positive r this has one solution, namely

$$u = \frac{(t+2a)+\sqrt{t}\sqrt{t+4a}}{2} = r^2(t), \qquad r(t) = \frac{1}{2}\Big(\sqrt{t}+\sqrt{t+4a}\Big)$$

The Jacobian is given by

$$\frac{du}{dt} = 2r(t)r'(t) = \frac{1}{4}\Big(\sqrt{t}+\sqrt{t+4a}\Big)\Big(\frac{1}{\sqrt{t}}+\frac{1}{\sqrt{t+4a}}\Big) = \frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}.$$

Recall that $U \sim \chi^2_k$ and let f_U denote its density, we have

$$f_T(t) = f_U(u(t))\Big|\frac{du}{dt}\Big| = \frac{1}{2^{k/2}\Gamma(k/2)}\Big(r^2(t)\Big)^{\frac{k}{2}-1}\exp\Big\{-\frac{r^2(t)}{2}\Big\}\frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}.$$

Combining yields

$$f_T(t) = \delta(t)F_{\chi^2_k}(a) + \frac{2^{-k/2}}{\Gamma(k/2)}\frac{r^k(t)}{\sqrt{t}\sqrt{t+4a}}\exp\Big\{-\frac{r^2(t)}{2}\Big\},$$

with

$$r(t) = \frac{1}{2}\Big(\sqrt{t}+\sqrt{t+4a}\Big),$$

as desired.

□

2.3.1. Some results related to the t-distribution

Theorem 2.1.2 showed that if X is spherically distributed with independent components then it follows a normal distribution. A natural next question would be to ask about other distributions, here we will consider the t-distribution often used in hypothesis testing.

Definition 2.3.4. A random variable X is said to follow a non-central multivariate t-distribution with ν degrees of freedom and non centrality parameter θ if X has the same law as

$$\frac{W+\theta}{\sqrt{U/\nu}}$$

where $W \sim N_k(0, I_k)$ and $U \sim \chi^2_\nu$.

Note that for general θ the multivariate t-distribution is actually not spherically distributed and this is only the case when $\theta = 0$. Recall that for spherically distributed random variables we can write $f_X(x) = g_X(||x-\theta||^2)$ so when $\theta = 0$ we may write

$$f_X(x) \propto \left(1+\frac{1}{\nu}x^T I_k^{-1} x\right)^{-\frac{1}{2}(\nu+k)} \implies g_X(u) \propto \left(1+\frac{u}{\nu}\right)^{-\frac{1}{2}(\nu+k)}.$$

The distribution of the norm of X when X follows a multivariate t-distribution can actually be found with a little effort. First we need one more definition.

Definition 2.3.5. A random variable X is said to follow a F-distribution with ν and k degrees of freedom if X has the same law as

$$\frac{U_1/\nu}{U_2/k}$$

where $U_1 \sim \chi^2_\nu$ and $U_2 \sim \chi^2_k$.

Recall that we can calculate the norm of a vector by taking the root of the dot product with itself.

Proposition 2.3.6. *Let X follow a k-variate (central) t-distribution with ν degrees of freedom and scale matrix I_k. Then $||X||^2/k$ follows an F-distribution with k and ν degrees of freedom.*

Proof. From Definition 2.3.4 we know we can write X as a ratio of a multivariate normal distribution and the root of a scaled chi-squared distribution. This implies that we have

$$\frac{X^T X}{k} = \frac{1}{k}\left(\frac{W}{\sqrt{U/\nu}}\right)^T\left(\frac{W}{\sqrt{U/\nu}}\right) = \frac{||W||^2/k}{U/\nu}.$$

Recall that $||W||^2 \sim \chi^2_k$ and thus by definition $||X||^2/k$ follows an F-distribution with k and ν degrees of freedom.

□

In the same setting as Proposition 2.3.6 the actual distribution of $R = ||X||$ follows through a simple transform, or more precisely. Let f be the density of an $F_{k,\nu}$-distribution. Then the density of R, denoted f_R is given by

$$f_R(r) = \frac{2r}{k} f\left(\frac{r^2}{k}; k, \nu\right)$$

Since the central multivariate t-distribution is actually spherically distributed it might make sense to use the positive part James-Stein test statistic.

Corollary 2.3.7. *Let $X \sim t_\nu(0, I_k)$. Then the positive part James-Stein test statistic has the following density*

$$f_T(t) = F_{k,\nu}\Big(\frac{a}{k}\Big)\delta(t) + \frac{1}{k} f_F\Big(\frac{r^2(t)}{k}; k, \nu\Big)\frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}\mathbb{1}_{\{t>0\}},$$

in which f_F is the density of an F-distribution with k, ν degrees of freedom.

Proof. Similar to as in Corollary 2.3.3 we split the two cases for $T(X)$. We know that $||X||^2/k \sim F_{k,\nu}$ from which it follows

$$\mathbb{P}\Big\{T(X) = 0\Big\} = \mathbb{P}\Big\{\frac{||X||^2}{k} \leq \frac{a}{k}\Big\} = F_{k,\nu}(a/k).$$

And for the same function g we get

$$u = r^2(t), \qquad r(t) = \frac{1}{2}\Big(\sqrt{t} + \sqrt{t+4a}\Big), \qquad \frac{du}{dt} = \frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}.$$

Let $U = ||X||^2$ then we have

$$f_U(u) = \frac{1}{k} f_F\Big(\frac{u}{k}; k, \nu\Big) \quad \text{and} \quad f_T(t) = f_U\big(r^2(t)\big)\Big|\frac{du}{dt}\Big| = \frac{1}{k} f_F\Big(\frac{r^2(t)}{k}; k, \nu\Big)\frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}.$$

Combining everything gives us

$$f_T(t) = F_{k,\nu}\Big(\frac{a}{k}\Big)\delta(t) + \frac{1}{k} f_F\Big(\frac{r^2(t)}{k}; k, \nu\Big)\frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}}\mathbb{1}_{\{t>0\}},$$

in which f_F is the density of an F-distribution with k, ν degrees of freedom.

□

The density function of an $F_{k,\nu}$-distribution actually has a closed form, it is given by

$$f_F(x) = \frac{\Gamma\big(\frac{k+\nu}{2}\big)}{\Gamma\big(\frac{k}{2}\big)\Gamma\big(\frac{\nu}{2}\big)}\Big(\frac{k}{\nu}\Big)^{k/2} x^{\frac{k}{2}-1}\Big(1 + \frac{k}{\nu}x\Big)^{-(k+\nu)/2}.$$

Plugging this in the continuous part of our density for when X follows a central t-distribution we get

$$\frac{1}{k}\Bigg(\frac{\Gamma\big(\frac{k+\nu}{2}\big)}{\Gamma\big(\frac{k}{2}\big)\Gamma\big(\frac{\nu}{2}\big)}\Big(\frac{k}{\nu}\Big)^{k/2}\Big(\frac{r^2(t)}{k}\Big)^{\frac{k}{2}-1}\Big(1 + \frac{r^2(t)}{\nu}\Big)^{-(k+\nu)/2}\Bigg)\frac{r^2(t)}{\sqrt{t}\sqrt{t+4a}},$$

which may be simplified to

$$\frac{\Gamma\big(\frac{k+\nu}{2}\big)\nu^{-k/2}}{\Gamma\big(\frac{k}{2}\big)\Gamma\big(\frac{\nu}{2}\big)}\frac{r^k(t)}{\sqrt{t}\sqrt{t+4a}}\Big(1 + \frac{r^2(t)}{\nu}\Big)^{-(k+\nu)/2}. \tag{2.3.2}$$

It is a well known fact that as $\nu \to \infty$ that we have that the law of t_ν converges to the law of a normal distribution, so it makes sense this expression in the limit converges to the one we found for the normal distribution. To show that this works we approximate the Gamma function using Stirling, which yields

$$\lim_{\nu\to\infty}\frac{\Gamma\big(\frac{k+\nu}{2}\big)\nu^{-k/2}}{\Gamma\big(\frac{k}{2}\big)\Gamma\big(\frac{\nu}{2}\big)} = \lim_{\nu\to\infty}\frac{\sqrt{\pi\big(k+\nu-2\big)}\big(\frac{k+\nu-2}{2e}\big)^{(k+\nu)/2-1}}{\Gamma\big(\frac{k}{2}\big)\sqrt{\pi\big(\nu-2\big)}\big(\frac{\nu-2}{2e}\big)^{\nu/2-1}\nu^{k/2}}.$$

Letting $\nu \to \infty$ is equivalent to letting $(\nu-2) \to \infty$ so after combining terms and substituting $\nu' = \nu-2$ we get

$$\frac{(2e)^{-k/2}}{\Gamma\left(\frac{k}{2}\right)} \lim_{\nu' \to \infty} \sqrt{1+\frac{k}{\nu'}}\left(1+\frac{k}{\nu'}\right)^{\frac{\nu'}{2}}\left(\frac{1+k/\nu'}{1+2/\nu'}\right)^{\frac{k}{2}}.$$

Now if we exponentiate and take the natural logarithm of this expression we can apply the Taylor series on the separate limits to get

$$\lim_{\nu' \to \infty} \frac{\nu'}{2} \ln\left(1+\frac{k}{\nu'}\right) = \frac{k}{2},$$
$$\lim_{\nu' \to \infty} \frac{k}{2}\left(\ln\left(1+\frac{k}{\nu'}\right) - \ln\left(1+\frac{2}{\nu'}\right)\right) = 0.$$

Substituting these results back into our original limit gives us

$$\lim_{\nu \to \infty} \frac{\Gamma\left(\frac{k+\nu}{2}\right)\nu^{-k/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} = \frac{1}{\Gamma\left(\frac{k}{2}\right)}\left(\frac{e}{2e}\right)^{\frac{k}{2}} = \frac{2^{-k/2}}{\Gamma(k/2)}$$

For the term that looks a lot like the definition of the exponential it turns out to be exactly that in the limit

$$\lim_{\nu \to \infty} \exp\left\{-\frac{k+\nu}{2} \ln\left(1+\frac{r^2(t)}{\nu}\right)\right\} = \lim_{\nu \to \infty} \exp\left\{-\frac{k+\nu}{2}\frac{r^2(t)}{\nu}\right\} = \exp\left\{-\frac{r^2(t)}{2}\right\}.$$

Combining all of this we get that Expression 2.3.2 is equal to

$$f_T(t) = \delta(t) F_{\chi^2_k}(a) + \frac{2^{-k/2}}{\Gamma(k/2)} \frac{r^k(t)}{\sqrt{t}\sqrt{t+4a}} \exp\left\{-\frac{r^2(t)}{2}\right\},$$

in the limit, exactly the expression we found for the central normal distribution.

3

Hypothesis testing

We now shift our focus to the subject of hypothesis testing using the positive part James-Stein test statistic. Recall that in hypothesis testing we are interested whether or not the parameter is with some probability not contained in a specified set. Whether or not the set contains the parameter is formulated as a *null-hypothesis* and *alternative-hypothesis*.

Let θ be the parameter of interest, Θ be the set of all possible values of θ and consider the two hypothesises

$$H_0 : \theta \in \Theta_0 \qquad \text{and} \qquad H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

Let K_T be the set of all value of $T(X)$ for which we reject the null-hypothesis, when testing at significance level α we want

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \Big\{ T(X) \in K_T \Big\} \le \alpha$$

This means we want the probability of us falsely rejecting the null-hypothesis to be at most α, often taken to be 0.05.

3.1. Behaviour of the law of $T(X)$ with respect to λ

Recall that from Proposition 2.1.5 we know that the distribution of the positive part James-Stein test statistic depends solely on θ through $\lambda = ||\theta||$. Intuitively, if the distance of the true mean were to increase with respect to the origin we would expect the density of the observed norm, R, to shift its mass away from the origin and thus it would be fair to assume that we have that the integral of f_R over some interval $(0, r)$ to be decreasing with respect to λ. Recall that when $X \sim N_k(\theta, I_k)$ that $R = ||X|| \sim \chi_k(\lambda)$.

Proposition 3.1.1. *Let $J \sim$ Poisson(μ) and $\{a_n\}_{n\ge 0}$ be a decreasing sequence. Then the Poisson mixture*

$$P(\mu) = \sum_{n=0}^{\infty} \mathbb{P}_\mu \big(J = n\big) a_n$$

is a decreasing function with respect to μ.

Proof. Taking the derivative of the infinite sum, which is valid since the sum converges, we get

$$\frac{d}{d\mu} P(\mu) = \sum_{n=0}^{\infty} \Big(\frac{n\mu^{n-1}}{n(n-1)!} - \frac{\mu^n}{n!} \Big) e^{-\mu} a_n = \sum_{n=0}^{\infty} \Big(\mathbb{P}\big(J = n-1\big) - \mathbb{P}\big(J = n\big) \Big) a_n.$$

With some index shifting, and applying that $\mathbb{P}(J=-1)=0$ we get

$$\sum_{n=0}^{\infty}\Big(\mathbb{P}\big(J=n-1\big)-\mathbb{P}\big(J=n\big)\Big)a_n=\sum_{n=0}^{\infty}\big(a_{n+1}-a_n\big)\mathbb{P}\big(J=n\big).$$

Since $\{a_n\}_{n\geq 0}$ is a decreasing sequence we have $(a_{n+1}-a_n)<0$ so this entire sum is negative and thus our Poisson mixture is decreasing with respect to μ.

□

Let $X,Y\sim N(0,1)$ then we have

$$\mathbb{P}\big(X^2+Y^2\leq t\big)=\frac{1}{2\pi}\int_{x^2+y^2\leq t}e^{-\frac{1}{2}(x^2+y^2)}\,dydx=\int_0^t\frac{1}{2}e^{-\frac{1}{2}s}ds.$$

So $X^2+Y^2\sim\mathrm{Exp}(\frac{1}{2})$. It can easily be shown, using characteristic functions for example, that the sum of exponential distributed random variables follows a Gamma distribution. I.e. Let $Z_i\sim\mathrm{Exp}(\frac{1}{2})$ then

$$\sum_{n=1}^{k}Z_i\sim\mathrm{Gamma}\big(k,\frac{1}{2}\big)$$

and thus we have that the chi-squared distribution with k degrees of freedom and Gamma distribution with parameters $\frac{k}{2}$ and $\frac{1}{2}$ are equivalent in distribution. From this follows a closed form expression for the cumulative density function of a chi-squared random variable from which it can be derived that this is a decreasing function with respect to k. Alternatively an intuitive argument, which can be made formal, is that when we increase the parameter k we are adding more strictly positive random variables and thus the probability of this distribution assuming a value smaller or equal to some fixed t decreases.

Lemma 3.1.2. *Let R follow a non-central chi distribution with k degrees of freedom, non-centrality parameter λ and $F_R(r;k,\lambda)$ its cumulative distribution function. Then $F_R(r;k,\lambda)$ is a decreasing function with respect to λ.*

Proof. We know $R^2\sim\chi_k^2(\lambda^2)$. Let $J\sim\mathrm{Poisson}(\lambda^2/2)$, the cumulative probability function of a non-central chi-squared distribution may be written as Poisson mixture as follows

$$F_{R^2}(r;k,\lambda)=\sum_{n=0}^{\infty}\mathbb{P}\big(J=n\big)F_{\chi^2}(r;k+2n)=e^{-\lambda^2/2}\sum_{n=0}^{\infty}\frac{(\lambda^2/2)^n}{n!}F_{\chi^2}(r;k+2n)$$

where $F_{\chi^2}(r;\nu)$ denotes the cumulative probability function of a central chi-squared distribution with ν degrees of freedom (Johnson et al., 1995). We know that $F_{\chi^2}(r;\nu)$ is a decreasing function with respect to ν. By increasing λ we are also increasing $\lambda^2/2$. Applying Proposition 3.1.1 then gives us that F_{R^2} is a decreasing function with respect to λ, the result for F_R follows from a simple transform.

□

In Chapter 2 it was shown that the density of $T(X)$ has a point mass, this point mass means that the cumulative function has a discontinuity at λ^2. Or more precisely, let $I(t;k,\lambda)$ denote the continuous part of the density of $T(X)$ then we have

$$F_T(t;k,\lambda)=F_R(\sqrt{a})\mathbb{1}_{\{t\geq\lambda^2\}}+\int_0^t I(s;k,\lambda)ds.$$

Furthermore, Lemma 3.1.2 showed that for at least the normal distribution the point mass portion of this cumulative function is decreasing with respect to λ so for the behaviour of the full density we are also interested in the behaviour of the continuous part.

Due to the complicated nature of the analytic density it is challenging to prove that this cumulative density is either increasing or decreasing with respect to λ. Numerical simulations suggest that F_T is a decreasing function with respect to λ but this remains an open problem.

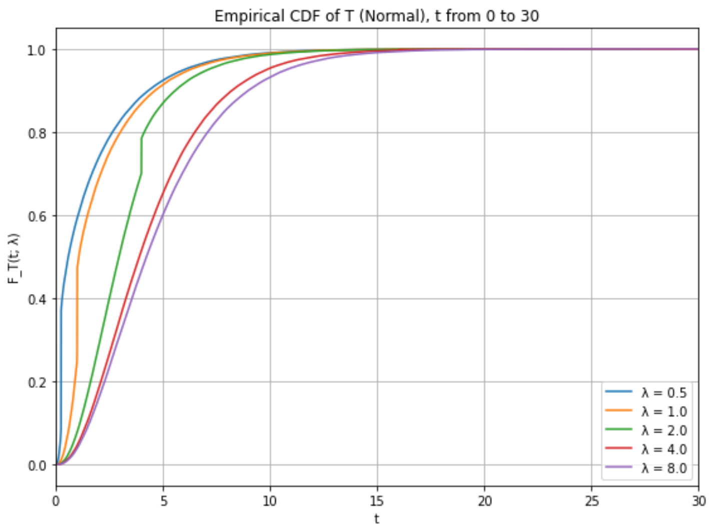


Figure 3.1: Empirical CDF of $T(X)$

To close this section we introduce one final proposition.

Proposition 3.1.3. *Let* $X \sim N_k(\theta, I_k)$. *Then* $\lim_{\lambda\to\infty} T_\lambda(X) \sim \chi_k^2$.

Proof. Write $X = Z + \theta$ where $Z \sim N_k(0, I_k)$ and set $S_\lambda = a/||X||^2$. As $\lambda \to \infty$ we have for all $\varepsilon > 0$ that $\mathbb{P}(a/||X||^2 < \varepsilon) = 1$. This in turn implies that we have $\mathbb{P}(1 \geq (1 - S_\lambda) > 1 - \varepsilon) = 1$ for all $\varepsilon > 0$. So for large λ we may write

$$T_\lambda(X) = \left|\left|(1 - S_\lambda)_+ X - \theta\right|\right|^2 = \left|\left|(1 - S_\lambda)(Z + \theta) - \theta\right|\right|^2 = \left|\left|(1 - S_\lambda)Z - S_\lambda\theta\right|\right|^2.$$

Since $(1 - S_\lambda) \xrightarrow{p} 1$ we have by Slutsky's lemma that $(1 - S_\lambda)Z \xrightarrow{d} Z$ and since$S_\lambda \xrightarrow{p} 0$ we have by the same lemma that $S_\lambda\theta \xrightarrow{p} 0$ and $(1 - S_\lambda)Z - S_\lambda\theta \xrightarrow{d} Z$. Finally, since the norm of a vector is a continuous mapping we conclude $T_\lambda(X) \xrightarrow{d} ||Z||^2 \sim \chi_k^2$.

□

3.2. Hypothesis testing

In classical hypothesis testing we want to test whether or not a parameter of interest, in this case θ, is contained in some set Θ_0 or its complement with respect to the full parameter space Θ. We call $\theta \in \Theta_0$ the null-hypothesis and $\theta \in \Theta_1 = \Theta / \Theta_0$ the alternative hypothesis. Then for some test statistic $T(X)$ and significance level α we are looking for a critical region K_T such that if $T(X) \in K_T$ we reject the null-hypothesis. Formally we require

$$\sup_{\theta \in \Theta_0} \mathbb{P}\Big(T(X) \in K_T\Big) \leq \alpha.$$

I.e. we want to find a set of values that $T(X)$ may attain such that if we reject the null hypothesis for those values there is at most a α probability of us falsely doing so.

As test statistic we will be using the positive part James-Stein test statistic which is given by

$$T(X) = \left|\left|\left(1 - \frac{a}{||X||^2}\right)_+ X - \theta\right|\right|^2,$$

and whose density we will denote f_T. We divide the problem of hypothesis testing into two subproblems. In the first case the set associated to the null-hypothesis only contains one value, in the second case this may be a general set.

3.2.1. Simple null-hypothesis

The term simple hypothesis testing refers to the case where Θ_0 is a singleton and thus the supremum no longer needed in the argument. More formally consider

$$H_0 : \theta = \theta_0 \qquad \text{and} \qquad H_1 : \theta \neq \theta_0$$

at significance level α. Let

$$F_T(t) = \int_0^t f_T(s)ds \qquad \text{and} \qquad q_T(\alpha) = \inf\{t : F_T(t) \geq \alpha\}$$

and since $F_T(t)$ is an increasing function we can write $K_T = [c_\alpha^2, \infty)$ for some $c_\alpha^2 \in \mathbb{R}_{\geq 0}$. If we let $c_\alpha^2 = q_T(1-\alpha)$ then we have

$$\mathbb{P}_{\theta_0}\Big\{T(X) \geq c_\alpha^2\Big\} \leq \alpha$$

as desired.

3.2.2. Connected region hypothesis testing

In the first part of this subsection we covered the somewhat trivial case of hypothesis testing with a simple null-hypothesis. We are now ready to cover the more complicated case where our set Θ_0 contains more than one value. Recall that in general hypothesis testing we split our parameter space Θ in two sets, one corresponding to the null-hypothesis, Θ_0, and the other to the alternative hypothesis, Θ_1. Let α denote our significance level which is the maximum probability we accept of a false rejection. Formally we have

$$H_0 : \theta \in \Theta_0 \qquad \text{and} \qquad H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

Let K_T be the set of all value of $T(X)$ for which we reject the null-hypothesis, when testing at significance level α we want

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\Big\{T(X) \in K_T\Big\} \leq \alpha.$$

To be able to calculate this we need to know for which value of $\theta \in \Theta_0$ this expression obtains its supremum. If we choose to make the assumption that F_T is a decreasing function with respect to λ this supremum is obtained at the $\theta \in \Theta_0$ for which we have the smallest norm.

3.2.3. Worked example

Let X follow a 4-dimensional normal distribution with unknown mean θ and known variance matrix I_4, the identity matrix. We want to test whether or not the true mean of this distribution is equal to the zero vector or not at significance level $\alpha = 0.05$. We formalise this as the two hypothesises

$$H_0 : \theta = 0 \qquad \text{and} \qquad H_1 : \theta \neq 0.$$

Under the null hypothesis the density of the positive part James-Stein test statistic was given by

$$f_T(t) = \delta(t) F_{\chi^2_4}(2) + \frac{(\sqrt{t} + \sqrt{t+8})^4}{64\sqrt{t}\sqrt{t+8}} \exp\left\{ -\frac{(\sqrt{t} + \sqrt{t+8})^2}{8} \right\},$$

in which we used $k = 4$ and $a = k - 2 = 2$. Numerically integrating this quantity gives us the following cumulative density function.

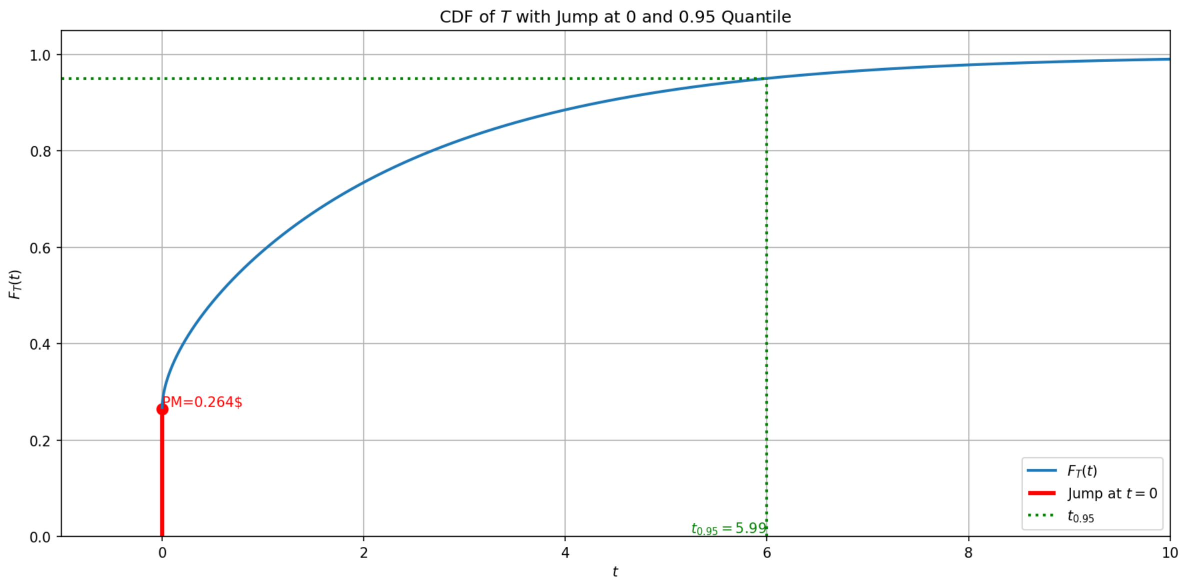


Figure 3.2: CDF of $T(X)$ when $X \sim N_4(0, I_4)$

The jump at $t = 0$ is the point mass part of our density which is located at the mean, we numerically find that the probability of $T(X)$ attaining a value less that roughly 6 is 0.95 and thus if $T(X)$ were to exceed this value we would reject the null-hypothesis.

Since we have $\theta = 0$ we may write the positive part James-Stein test statistic as

$$T_0(X) = \left(1 - \frac{2}{||X||^2}\right)_+^2 ||X||^2.$$

We have that this test statistic is less than or equal to the previously found quantile of roughly 6 when $||X|| \leq 3.094$ which makes the volume of the critical region for X that of a 4-dimensional sphere with radius 3.094. When using the standard chi-squared test we would reject H_0 if we have $||X||^2 > 9.488$ which is equivalent to when $||X|| > 3.080$. Since the distribution of X is the same in both cases it is expected that the volume of the critical regions for X are the same .

4

Constructing Confidence Sets

In the classic setting we construct an analytical confidence set using a pivot, a statistic that is a function of our desired parameter but whose distribution does not depend on said parameter. Common examples of such pivots are the Z test statistic for a one-dimensional normal distribution. We have

$$Z = \sqrt{n}\frac{X-\theta}{\sigma} \sim N(0,1)$$

which may then be used to construct a confidence interval of any desired level as follows

$$\mathbb{P}\left\{\xi_{\alpha/2} \leq \sqrt{n}\frac{X-\theta}{\sigma} \leq \xi_{1-\alpha/2}\right\} = \mathbb{P}\left\{X - \xi_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq X + \xi_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

in which we used the symmetry of the standard normal distribution. This is also where we run into our first issue. The distribution of our positive part James-Stein test statistic depends on the parameter we are trying to construct an interval for through $\lambda = ||\theta||$ so we need a different approach.

4.1. The Quantiles of the positive part James-Stein test statistic

Before we look at techniques to approximate the quantiles using estimators for $\lambda = ||\theta||$ we consider the, unrealistic, case where λ is known. When λ is known we can calculate the true quantiles of the positive part James-Stein test statistic, they are approximately given by

$\lambda \backslash k$	3	4	5	6	7	8	9	10
1.0	6.10	6.16	6.25	6.31	6.34	6.37	6.40	6.43
2.0	6.43	6.79	7.09	7.33	7.51	7.66	7.78	7.90
3.0	7.45	8.35	8.98	9.01	9.04	9.34	9.61	9.85
4.0	7.66	8.89	9.88	10.69	11.38	11.98	12.49	12.94
5.0	7.72	9.10	10.30	11.32	12.25	13.06	13.78	14.44
6.0	7.75	9.22	10.51	11.68	12.76	13.72	14.62	15.47
7.0	7.78	9.28	10.66	11.92	13.09	14.17	15.20	16.13
8.0	7.78	9.34	10.75	12.07	13.30	14.47	15.56	16.61
9.0	7.81	9.37	10.84	12.19	13.45	14.68	15.83	16.94
10.0	7.81	9.40	10.87	12.25	13.57	14.83	16.04	17.18

Table 4.1: True, approximated, 0.95 Quantiles (λ, k)

These, and the following, values were obtained using python code provided as an appendix. In the usual case when $X \sim N_k(\theta, I_k)$ we know that $||X-\theta||^2 \sim \chi^2_k$ and thus we can construct a confidence set for θ using the $(1-\alpha)$-quantiles of the χ^2_k distribution. The corresponding quantiles can be approximated and are equal to

k	3	4	5	6	7	8	9	10
$\chi^2_{k,0.95}$	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Comparing these values we can see that in all cases the quantiles of the positive part James-Stein test statistic are smaller.

4.2. Estimation of λ

Now that we have shown that when we know λ we can calculate the quantiles of the positive part James-Stein test statistic. We now consider some techniques of estimating these quantiles when λ is not known. Note that due to computational constraints a small sample size was used.

4.2.1. Naïve MLE

First we consider the maximum likelihood estimator for θ from which we can estimate $\lambda = ||\theta||$. The maximum likelihood estimator is well known and given by $\hat{\theta}_{MLE} = X$. Though it might be tempting to say that $||\hat{\theta}_{MLE}||$ is an unbiased estimator for λ this is actually not the case. $||\hat{\theta}_{MLE}||$ actually always over estimates the parameter λ. To see this note that we have $X = \theta + Z$ where $Z \sim N_k(0, I_k)$ and define the function $f : \mathbb{R}^k \to \mathbb{R}$ by $f(u) = ||u||$. Note that the Euclidean norm is a convex function and thus by Jensen's inequality we have

$$\mathbb{E}\big(||X||\big) = \mathbb{E}\big(f(\theta + Z)\big) \geq f\big(\mathbb{E}(\theta + Z)\big) = ||\theta + \mathbb{E}(Z)|| = ||\theta||,$$

and since Z non-degenerate and thus $\mathbb{P}(Z = 0) = 0$ we have $\mathbb{E}(||X||) > \lambda$. When using the naïve MLE we get the following quantiles.

$\lambda \backslash k$	3	4	5	6	7	8	9	10	Average
1.0	+0.41	+0.74	+0.72	+2.43	+2.16	+2.99	+4.88	+5.31	2.46
2.0	+0.52	+0.93	+0.82	+1.97	+3.10	+3.33	+2.57	+5.00	2.28
3.0	+0.10	+0.37	+0.71	+1.03	+1.78	+2.24	+2.86	+3.89	1.62
4.0	-0.40	-0.17	+0.40	+0.31	+0.78	+0.98	-0.06	+1.04	0.52
5.0	+0.04	+0.09	+0.21	+0.00	+0.30	-0.03	+0.24	+0.37	0.16
6.0	-0.01	+0.05	+0.14	+0.12	+0.24	-0.16	+0.32	+0.19	0.15
7.0	-0.02	-0.02	+0.05	-0.02	+0.19	+0.16	+0.22	+0.42	0.14
8.0	+0.00	-0.01	+0.02	-0.02	+0.14	+0.26	-0.06	+0.00	0.06
9.0	+0.00	-0.01	-0.01	+0.00	-0.05	-0.06	+0.11	+0.03	0.03
10.0	+0.00	-0.01	-0.02	-0.04	-0.03	+0.03	+0.12	+0.16	0.05
Average	0.15	0.24	0.31	0.59	0.88	1.02	1.14	1.64	0.75

Table 4.2: Difference between true quantiles and estimated quantiles using the naïve MLE together with their absolute error and averages.

4.2.2. An unbiased estimator for λ^2

We have shown that the norm of the most likelihood estimator turns out to be positively biased so a reasonable next step would be to consider an unbiased estimator for λ. It turns out that in the non-trivial case finding such an estimator is quite hard but there does exist an easy unbiased estimator for λ^2. Consider $\hat{\lambda^2} = ||X||^2 - k$, we know $||X||^2 \sim \chi^2_k(\lambda^2)$

$$\mathbb{E}(||X||^2 - k) = \mathbb{E}(||X||^2) - k = \lambda^2 + k - k = \lambda^2$$

and thus a reasonable estimator for λ might be $\hat{\lambda}_{UB} = \sqrt{(||X||^2 - k)_+}$. When using this estimator we get the following quantiles.

$\lambda \backslash k$	3	4	5	6	7	8	9	10	Average
1.0	+0.95	+1.67	+2.02	+1.27	+1.05	+2.24	+2.92	+1.52	1.71
2.0	+0.07	+0.54	+0.00	+1.18	+1.43	+0.32	+0.56	+0.86	0.62
3.0	-0.21	-0.23	-0.44	+0.22	-0.55	-0.65	+0.10	+1.04	0.43
4.0	-0.59	-0.41	+0.17	-0.50	-0.08	-0.26	-1.20	-1.02	0.53
5.0	+0.03	+0.05	+0.11	-0.39	-0.13	-1.22	-1.11	-1.09	0.52
6.0	-0.03	+0.03	+0.08	-0.02	+0.03	-0.70	-0.24	-0.62	0.22
7.0	-0.02	-0.05	+0.02	-0.12	+0.10	-0.04	-0.03	+0.07	0.06
8.0	+0.00	-0.01	+0.01	-0.07	+0.06	+0.18	-0.30	-0.29	0.11
9.0	+0.00	-0.02	-0.02	-0.03	-0.12	-0.16	+0.00	-0.17	0.07
10.0	+0.00	-0.02	-0.05	-0.07	-0.07	-0.03	+0.04	+0.06	0.04
Average	0.19	0.30	0.29	0.39	0.36	0.58	0.65	0.67	0.43

Table 4.3: Difference between true quantiles and estimated quantiles using the unbiased estimator for λ^2 together with their absolute error and averages.

4.2.3. Positive part James-Stein estimator

A final obvious contestant is the norm of the positive part James-Stein estimator, i.e.

$$\hat{\lambda}^+_{JS} = \left|\left|\left(1 - \frac{a}{||X||^2}\right)_+ X\right|\right| = \left(||X|| - \frac{a}{||X||}\right)\mathbb{1}_{\{||X||^2 > a\}}$$

When using this estimator we get the following quantiles.

$\lambda \backslash k$	3	4	5	6	7	8	9	10	Average
1.0	+0.12	+1.16	+0.80	+1.20	+1.19	+1.62	+1.41	+0.55	1.01
2.0	+0.24	+0.15	-0.42	+0.48	+0.63	-0.06	-0.28	+0.05	0.29
3.0	-0.09	-0.11	-0.48	-0.46	-0.80	-1.09	-0.66	-0.44	0.52
4.0	-0.52	-0.37	+0.14	-0.66	-0.37	-0.89	-2.24	-1.93	0.89
5.0	+0.03	+0.05	+0.09	-0.52	-0.30	-1.61	-1.53	-2.32	0.81
6.0	-0.02	+0.03	+0.08	-0.06	-0.06	-0.98	-0.55	-1.10	0.36
7.0	-0.02	-0.05	+0.01	-0.15	+0.05	-0.14	-0.19	-0.13	0.09
8.0	+0.00	-0.01	+0.01	-0.08	+0.04	+0.13	-0.44	-0.48	0.15
9.0	+0.00	-0.02	-0.02	-0.05	-0.15	-0.22	-0.06	-0.29	0.10
10.0	+0.00	-0.02	-0.05	-0.09	-0.10	-0.06	+0.02	+0.01	0.04
Average	0.10	0.20	0.21	0.38	0.37	0.68	0.74	0.73	0.43

Table 4.4: Difference between true quantiles and estimated quantiles using the positive part James-Stein estimator together with their absolute error and averages.

4.2.4. Performance

To compare the three estimators for λ we computed, for each pair (k, λ), the average absolute error in the 95th-percentile of the test statistic centred around the positive part James-Stein test statistic. The naive maximum likelihood estimator appears to systematically overshoot when λ is small and becomes less biased for larger λ. This can intuitively explained by the steep derivative of the square root function for small values and the almost linear behaviour for large values. The unbiased λ^2 estimator tends to overshoot for very small λ, then mildly underestimates at moderate λ, before approaching negligible bias once λ is large. In contrast, $\hat{\lambda}^+_{JS}$ pulls the estimate closer to zero enough to reduce the overshoot of the naive most likelihood estimator when λ small, yet does not introduce severe undershoot at moderate λ. Overall, the James–Stein approach achieves the lowest average absolute error, closely followed by the unbiased-λ^2 method, whereas the naive most likelihood estimators error is larger larger unless $\|\theta\|$ is known to be large. Consequently, if one seeks to minimize average radius-error, and thus mis-coverage, across a range of (k,λ), the positive-part James–Stein estimator is recommended. The unbiased λ^2 estimator is a viable alternative once $\|\theta\|$ is not extremely small, and the naive most likelihood estimator should only be used for settings in which $\|\theta\|$ is known to be large.

4.3. Hypothesis testing Inversion

Another approach is to in some sense inverse the hypothesis testing process. This idea is based on theory from a textbook by Schervisch.

Proposition 4.3.1. *(Schervish, 2012) Let $g:\Omega\to G$ be a function.*
For each $y\in G$, let ϕ_y be a level-α non-randomised test of $H_0: g(\theta)=y$. Let $R(x)=\{y:\phi_y(x)=0\}$. Then R is a level $(1-\alpha)$ confidence set for $g(\theta)$.

- *Let R be a level-$(1-\alpha)$ confidence set for $g(\theta)$. For each $y\in G$ define*

$$\phi_y(x)=\mathbb{I}_{\{y\notin R(x)\}}.$$

Then, for each y,ϕ_y has level α as a test of $H_0: g(\theta)=y$.

In our case we take $g(\theta)=\theta$ and $G=\Theta=\mathbb{R}^k$. Thus, for each fixed candidate point $y\in\mathbb{R}^k$, we wish to test $H_0:\theta=y$ at level α. Under the null hypothesis we have $X\sim N_k(y,I_k)$. Let $T_y(X)$ denote the positive part James-Stein test statistic under this mean and define

$$c_{1-\alpha}(||y||)=\inf\{t:\mathbb{P}_y\big(T_y(X)\le t\big)\ge 1-\alpha\}$$

This gives us

$$\phi_y(x)=\begin{cases}1, & T_y(X)>c_{1-\alpha}(||y||),\\ 0, & T_y(X)\le c_{1-\alpha}(||y||).\end{cases}$$

Now Proposition 4.3.1 states that the set

$$R(X)=\{y\in\mathbb{R}^k:\phi_y(X)=0\}$$

which may be written as

$$R(X)=\{y\in\mathbb{R}^k: T_y(X)\le c_{1-\alpha}(||y||)\}$$

is a $(1-\alpha)$-confidence set for θ.

5
Conclusion

We set out to investigate the properties of test statistics centred around the positive part James-Stein estimator and potential applications within the landscape of hypothesis testing and the creation of confidence sets.

In Chapter 2 Theorem 2.1.2 showed that if we are working with a k-variate spherically distributed random variable around θ with independent components then it follows a normal distribution. Additionally formal definitions of the positive part James-Stein estimator and test statistic were given and their properties were studied. Subsection 2.2 found closed form expressions for the distribution of the radius of some spherically distributed random variable and Subsection 2.3 gave the density for the positive part James-Stein test statistic. The chapter also states the normal distribution case as a separate result and shows the link between the central t-distribution and normal distribution.

Chapter 3 sets out to apply the distribution found in Theorem 2.3.2 for simple and composite hypothesis testing. The behaviour of the positive part James-Stein test statistic was studied in Subsection 3.1 we showed that the point mass part of the density of $T(X)$ is decreasing with respect to λ. It it conjectured that the cumulative density of $T(X)$ is a decreasing function with respect to λ for which numerical motivation is provided. The subsection ends by showing that the distribution of $T(X)$ converges to that of a chi-squared distribution when λ tends to infinity. Subsection 3.2 goes over the process of applying these concepts to hypothesis tests and it is shown that at least in the 4-dimensional normal distribution case there is no advantage to using the positive part James-Stein test statistic over the classic chi-squared test statistic.

Finally Chapter 4 briefly goes over the construction of confidence sets using the positive part James-Stein test statistic. Section 4.1 approximates the true quantiles of the positive part James-Stein test statistic and compares them with the quantiles of a chi-squared distribution from which it is concluded that the positive part James-Stein test statistic allows for smaller radii. Section 4.2 uses a variety of estimators for λ to estimate these quantiles based on simulated samples. The subsection closes with a short conclusion in which is it concluded that the best estimator is based on the unbiased estimator for λ^2. Subsection 4.3 briefly goes over how one can apply hypothesis testing inversion to construct confidence sets using the positive part James-Stein test statistic.

6

Discussion and Further Research

Chapter 2 introduces a closed form expression for the law of the positive part James-Stein test statistic but in its current form it is not very practical. Perhaps utilising the fact that the boundaries of the integral in Theorem 2.3.2 are chosen such that $\nu \in [-1, 1]$ would allow for a cleaner formula. The chapter additionally places a large focus on the multivariate normal distribution but the class of spherical distributions is of course significantly larger.

Chapter 3 studies the behaviour of the previously found law but fails to make any formal conclusions about under what conditions this law attains its maximum. It also briefly considers hypothesis testing but there is still a lot to be said about the general case where the null-hypothesis might not be a singleton or neatly defined convex set.

In Chapter 4 it was shown that when λ is known that we can improve on the classic confidence set but in practice this is not the case. The chapter attempts to mediate this by introducing several estimators for λ but due to small sample size and computational constraints there is still work to be done here.

A

Derivation of the Beta integral

We want to find a closed form expression for

$$\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}e^{r\lambda t}dt = \int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}\sum_{n=0}^{\infty}\frac{(r\lambda t)^n}{n!}dt = \sum_{n=0}^{\infty}\frac{(r\lambda)^n}{n!}\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}t^n dt$$

in which the swapping of the order of integration and summation is justified by dominated convergence theorem. Let $u = t^2$, then

$$\frac{du}{dt} = 2t \implies \int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}t^n dt = \int_{0}^{1}(1-u)^{\frac{k-3}{2}}u^{\frac{n-1}{2}}du = B\Big(\frac{k-1}{2},\frac{n+1}{2}\Big),$$

in which $B(z_1, z_2)$ denotes the Beta function. Additionally note that when n is odd the original integral in the summation is the product of an even and odd function and is therefore equal to zero which implies we may write

$$\sum_{n=0}^{\infty}\frac{(r\lambda)^{2n}}{(2n)!}\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}t^{2n}dt = \frac{1}{2}\sum_{n=0}^{\infty}\frac{(r\lambda)^{2n}}{(2n)!}B\Big(\frac{k-1}{2},n+\frac{1}{2}\Big).$$

Recalling the identities for the Beta and modified Bessel functions

$$B\Big(\frac{k-1}{2},n+\frac{1}{2}\Big) = \frac{\Gamma\big(\frac{k-1}{2}\big)\Gamma\big(n+\frac{1}{2}\big)}{\Gamma\big(n+\frac{k}{2}\big)}, \qquad I_{\frac{k}{2}-1}(r\lambda) = \sum_{n=0}^{\infty}\frac{1}{n!\Gamma\big(n+\frac{k}{2}\big)}\Big(\frac{r\lambda}{2}\Big)^{2n+\left(\frac{k}{2}-1\right)}$$

we can write

$$\Gamma\big((k-1)/2\big)\sum_{n=0}^{\infty}\frac{(r\lambda)^{2n}}{(2n)!}\frac{\Gamma\big(n+\frac{1}{2}\big)}{\Gamma\big(n+\frac{k}{2}\big)} = \Gamma\big((k-1)/2\big)\sum_{n=0}^{\infty}\frac{(r\lambda)^{2n}}{(2n)!}\frac{\sqrt{\pi}(2n)!}{4^n n!\Gamma\big(n+\frac{k}{2}\big)}$$

in which we used $\Gamma(n+\frac{1}{2}) = \sqrt{\pi}(2n)!/(4^n n!)$. Collecting terms we get

$$\sqrt{\pi}\Gamma\big((k-1)/2\big)\sum_{n=0}^{\infty}\frac{(r\lambda)^{2n}}{4^n n!\Gamma\big(n+\frac{k}{2}\big)} = \frac{\sqrt{\pi}\Gamma\big((k-1)/2\big)}{\big(\frac{1}{2}r\lambda\big)^{k/2-1}}\sum_{n=0}^{\infty}\frac{1}{n!\Gamma\big(n+\frac{k}{2}\big)}\Big(\frac{r\lambda}{2}\Big)^{2n+\left(\frac{k}{2}-1\right)}$$

and thus finally we have

$$\int_{-1}^{1}(1-t^2)^{\frac{k-3}{2}}e^{r\lambda t}dt = \frac{\sqrt{\pi}\Gamma\big((k-1)/2\big)}{\big(\frac{1}{2}r\lambda\big)^{k/2-1}}I_{\frac{k}{2}-1}(r\lambda).$$

which was what we wanted to show.

B

Proof of Theorem 2.3.2

Proof. We consider the two cases separately, Proposition 2.3.1 covers the case where we have $||X||^2 \leq a$ so we now consider its complement, $||X||^2 > a$. Define $A(r) = r - a/r, R = ||X||$ and $Y = X/||X||$. We may write

$$T(X) = \Big|\Big|X - a\frac{X}{||X||^2} - \theta\Big|\Big|^2 = \Big|\Big|A(R)Y - \theta\Big|\Big|^2 = A^2(R)||Y||^2 + \lambda^2 - 2A(R)\langle Y, \theta\rangle.$$

Y lies on the unit sphere and thus its length is always one. Let $V = \langle Y, \theta/\lambda\rangle$ then we can write

$$T(X) = A^2(R) + \lambda^2 - 2A(R)\lambda V = m - bV$$

which implies that we have

$$\mathbb{P}\Big\{T(X) \leq t \Big| R = r\Big\} = \mathbb{P}\Big\{V \geq \frac{m-t}{b}\Big| R = r\Big\} = 1 - \mathbb{P}\Big\{V \leq \frac{m-t}{b}\Big| R = r\Big\}$$

The distribution of $T|R$ follows from Lemmas 2.2.2 and 2.2.1 which gives us

$$f_{T|R}(t|r) = \frac{1}{2A(r)\lambda} f_{V|R}\Big(\frac{A^2(r) + \lambda^2 - t}{2A(r)\lambda}\Big| r\Big)$$

which then gives us

$$f_T(t) = \Big(\int_0^{\sqrt{a}} f_R(r)dr\Big)\delta\big(t - ||\theta||^2\big) + \int_{\sqrt{a}}^{\infty} \frac{1}{2A(r)||\theta||} f_{V|R}\Big(\frac{A^2(r) + ||\theta||^2 - t}{2A(r)||\theta||}\Big| r\Big) f_R(r)dr.$$

Recall that since we have $v \in [-1,1]$ we can actually swap the bounds of the integral for finite values. We require that

$$-1 \leq \frac{A^2(r) + \lambda^2 - t}{2A(r)\lambda} \leq 1 \implies \big(A(r) - \lambda\big)^2 \leq t \leq \big(A(r) + \lambda\big)^2.$$

Working out these inequalities we get the bounds

$$r_-(t,\lambda,a) = \frac{|\lambda - \sqrt{t}| + \sqrt{|\lambda - \sqrt{t}|^2 + 4a}}{2},$$

$$r_+(t,\lambda,a) = \frac{(\lambda + \sqrt{t}) + \sqrt{(\lambda + \sqrt{t})^2 + 4a}}{2}$$

Finally, combining the densities we get

$$f_T(t) = \Big(\int_0^{\sqrt{a}} f_R(s)ds\Big)\delta(t - \lambda^2) + \int_{r_-}^{r_+} \frac{f_{V,R}(v(r), r)}{2A(r)\lambda} dr, \quad v(r) = \frac{A^2(r) + \lambda^2 - t}{2A(r)\lambda}$$

as desired.

□

C
Python code

```
import numpy as np
import pandas as pd
from scipy.stats import ncx2, chi2
from scipy.special import gamma
from scipy.integrate import quad


def theoretical_cdf_T(lam, k=5, a=None, t_min=0.0, t_max=30.0, n_points
    =300):
    """
    Compute the theoretical CDF of
        T = ||X_shrink      ||^2
    when X ~ N_k(  , I_k), ||  || = lam, and a = k     2 (shrinkage
        factor).

    Special case: if lam == 0, then X ~ N_k(0, I_k) and one obtains the
        closedform
    density
        f_T(t) =   (t) * P(  _k       a)
               + (2^{-k/2}/  (k/2)) * [r(t)^k / (  t      (t + 4a))] *
                   exp(  r  (t)  /2),
    with r(t) = [  t  +    (t + 4a ) ] / 2.

    Returns
    -------
    t_grid : ndarray, shape (n_points,)
    F_theor: ndarray, shape (n_points,)
        Theoretical CDF values F_T(t; lam) evaluated at every point in
            t_grid.
    """
    if a is None:
        a = k - 2

    # 1) Build a uniform grid of t-values
    t_grid = np.linspace(t_min, t_max, n_points)
    dt = t_grid[1] - t_grid[0]

    # -----------------------------------------------------------
    # Special case: lam == 0
```

```
    # -----------------------------------------------------------------
    if lam == 0:
        # (a) The point mass at t = 0 is P( ||X||   a ) for X ~ N_k(0,I)  <=>   _k   a
        p_atom = chi2.cdf(a, df=k)

        # (b) The continuous density for t > 0 is
        #       f_cont(t) = [ 2^{-k/2} /  (k/2) ] * [ r(t)^k / (  t  (t + 4a) ) ] * exp(  r(t)^2 / 2 ),
        #    where r(t) = [  t  +   (t + 4a ) ] / 2.
        #
        #    We will fill f_cont_vals[i] only for t_grid[i] > 0.  At t = 0, set f_cont = 0,
        #    since the  delta mass at 0 is handled by p_atom.

        # Precompute the constant 2^{-k/2} /  (k/2)
        prefactor = (2.0 ** ( -k/2 )) / gamma(k/2)

        f_cont_vals = np.zeros_like(t_grid)
        for i, t in enumerate(t_grid):
            if t <= 0:
                # At exactly t = 0, the continuous part vanishes; the delta mass handles it.
                f_cont_vals[i] = 0.0
            else:
                # compute r(t)
                sqrt_t   = np.sqrt(t)
                sqrt_tp4a = np.sqrt(t + 4.0 * a)
                r_t = 0.5 * (sqrt_t + sqrt_tp4a)

                # density at t > 0
                #   f_cont(t) = prefactor * [ r(t)^k / ( sqrt(t) * sqrt(t + 4a) ) ] * exp(-r(t)^2 / 2)
                f_cont_vals[i] = (
                    prefactor
                    * (r_t ** k)
                    / ( sqrt_t * sqrt_tp4a )
                    * np.exp(-0.5 * (r_t ** 2))
                )

        # (c) Build continuous CDF by Riemann sum (left Riemann / trapezoid omitted, but dt is small)
        F_cont = np.cumsum(f_cont_vals) * dt

        # (d) Add the point mass at t = 0 for all t >= 0
        F_theor = F_cont.copy()
        idx0 = np.searchsorted(t_grid, 0.0, side='left')
        if idx0 < len(F_theor):
            F_theor[idx0:] += p_atom

        return t_grid, F_theor

    # -------------------------------------------------------
    # General case: lam > 0
    # -------------------------------------------------------

```

```
    # (1) Point mass at t = lam^2 from R^2    ||X||  ~   _nc  (k,
        lam )
    p_atom = ncx2.cdf(a, df=k, nc=lam**2)

    # (2) Continuous density prefactor
    prefac_denominator = np.sqrt(np.pi) * gamma((k - 1) / 2) * (2 ** (k
        / 2))
    C_k = np.exp(-(lam**2 + a) / 2) / (prefac_denominator * lam)

    # (3) Loop over t_grid, integrate from  r   to  r
    f_cont_vals = np.zeros_like(t_grid)
    for i, t in enumerate(t_grid):
        sqrt_t = np.sqrt(t)
        diff   = lam - sqrt_t

        #  r  (t) = [ |        t  | +    { (         t  )   + 4a } ] / 2
        #  r  (t) = [ (   +   t ) +    { (   +   t )   + 4a } ] / 2
        r_minus = (np.abs(diff) + np.sqrt(diff**2 + 4.0 * a)) / 2.0
        r_plus  = ((lam + sqrt_t) + np.sqrt((lam + sqrt_t)**2 + 4.0 * a
            )) / 2.0

        if r_minus >= r_plus:
            f_cont_vals[i] = 0.0
            continue

        def integrand(r):
            # v = [ (r       a)        (t           ) r   ] / [ 2 r (r
                    a)    ]
            num_v   = (r**2 - a)**2 - (t - lam**2) * (r**2)
            denom_v = 2.0 * r * (r**2 - a) * lam
            v = num_v / denom_v

            base = (1.0 - v**2)
            if base < 0:
                base = 0.0

            exponent = (r**2 * (lam**2 - t)) / (2.0 * (r**2 - a))
            return (r**k / (r**2 - a)) * np.exp(exponent) * (base ** ((
                k - 3) / 2))

        integral_val, _ = quad(integrand, r_minus, r_plus, epsabs=1e-9,
            epsrel=1e-9)
        f_cont_vals[i] = C_k * integral_val

    # (4) Build continuous CDF by Riemann sum
    F_cont = np.cumsum(f_cont_vals) * dt

    # (5) Add the jump of size p_atom at t = lam^2
    F_theor = F_cont.copy()
    idx_jump = np.searchsorted(t_grid, lam**2, side='right')
    if idx_jump < len(F_theor):
        F_theor[idx_jump:] += p_atom

    return t_grid, F_theor
```

```
# --- Compute only the 0.95 quantile of T for each (lam, k) pair ----

lam_list = np.linspace(1, 10, 10)   # 10 values from 1 to 10
k_list   = np.linspace(3, 10, 8)    # 8 values from 3 to 10
N        = 5


#  P r e allocate  a 2D array for the 0.95  quantiles
quantile_T = np.zeros((len(lam_list), len(k_list)))

# Use the first (lam, k) pair to get the t_grid length
t_vals_sample, _ = theoretical_cdf_T(lam_list[0], k=int(k_list[0]), a=int(k_list[0]) - 2,
                                     t_min=0, t_max=100, n_points=1000)

for i, lam in enumerate(lam_list):
    print(f"finding theoretical quantiles for lambda = {lam}")
    for j, k_val in enumerate(k_list):
        k_int = int(k_val)
        a_int = k_int - 2

        # Compute the full theoretical CDF on the grid
        t_vals, F_vals = theoretical_cdf_T(lam,
                                           k=k_int,
                                           a=a_int,
                                           t_min=0,
                                           t_max=30,
                                           n_points=len(t_vals_sample))

        # Find the smallest t such that F_T(t) >= 0.95
        # If no value reaches 0.95, assign NaN
        if np.any(F_vals >= 0.95):
            idx_095 = np.argmax(F_vals >= 0.95)
            quantile_T[i, j] = t_vals[idx_095]
        else:
            quantile_T[i, j] = np.nan

# Now quantile_T[i, j] holds the 0.95  quantile  for lam_list[i], k_list[j].


# Create DataFrame with   as rows and k as columns
df_quantiles = pd.DataFrame(
    quantile_T,
    index=np.round(lam_list, 2),
    columns=np.round(k_list, 2)
)
df_quantiles.index.name = '  '
df_quantiles.columns.name = 'k'

print(df_quantiles.round(2))


quantiles_naive          = np.zeros((len(lam_list), len(k_list)))
quantiles_unbiased       = np.zeros((len(lam_list), len(k_list)))
```

```
quantiles_stein          = np.zeros((len(lam_list), len(k_list)))



for i, lam in enumerate(lam_list):
    print(f"Finding estimated quantiles for lambda = {lam}")
    for j, k_val in enumerate(k_list):
        k_int = int(k_val)
        a_int = k_int - 2

        # Construct   vector of norm   along first coordinate
        theta = np.zeros(k_int)
        theta[0] = lam

        # Simulate N samples X ~ N_k(  , I_k)
        X = np.random.normal(size=(N, k_int)) + theta

        # Compute norms ||X||
        norms = np.linalg.norm(X, axis=1)

        # Estimator 1: Naive MLE ||X||
        est_naive_lambda = norms
        q_naive  = np.zeros(N)

        # Estimator 2: Unbiased for  ^2: sqrt((||X||^2 - k)_+)
        est_ub_lambda = np.sqrt(np.maximum(norms**2 - k_int, 0))
        q_unbiased = np.zeros(N)


        # Estimator 3: Positive part James-Stein: (||X|| - a/||X||) I{||X||^2 > a}
        est_js_lambda = np.where(norms**2 > a_int, norms - a_int / norms, 0)
        q_stein  = np.zeros(N)


        for I in range(N):
            t_vals_1, F_vals_1 = theoretical_cdf_T(est_naive_lambda[I],
                                                   k=k_int,
                                                   a=a_int,
                                                   t_min=0,
                                                   t_max=30,
                                                   n_points=len(t_vals_sample))

            t_vals_2, F_vals_2 = theoretical_cdf_T(est_ub_lambda[I],
                                                   k=k_int,
                                                   a=a_int,
                                                   t_min=0,
                                                   t_max=30,
                                                   n_points=len(t_vals_sample))

            t_vals_3, F_vals_3 = theoretical_cdf_T(est_js_lambda[I],
                                                   k=k_int,
                                                   a=a_int,
```

```
                                                                t_min=0,
                                                                t_max=30,
                                                                n_points=len(
                                                                    t_vals_sample))

                # Find the smallest t such that F_T(t) >= 0.95
                # If no value reaches 0.95, assign NaN
                if np.any(F_vals_1 >= 0.95):
                    idx_095 = np.argmax(F_vals_1 >= 0.95)
                    q_naive[I] = t_vals_1[idx_095]
                else:
                    print(F_vals_1[-1])
                    q_naive[I] = np.nan


                if np.any(F_vals_2 >= 0.95):
                    idx_095 = np.argmax(F_vals_2 >= 0.95)
                    q_unbiased[I] = t_vals_2[idx_095]
                else:
                    print(F_vals_2[-1])
                    q_unbiased[I] = np.nan

                if np.any(F_vals_3 >= 0.95):
                    idx_095 = np.argmax(F_vals_3 >= 0.95)
                    q_stein[I] = t_vals_3[idx_095]
                else:
                    print(F_vals_3[-1])
                    q_stein[I] = np.nan

                quantiles_naive[i, j]       = np.mean(q_naive)
                quantiles_unbiased[i, j]    = np.mean(q_unbiased)
                quantiles_stein[i, j]       = np.mean(q_stein)

print((quantiles_naive - quantile_T).round(2))
print(quantiles_unbiased.round(2))
print(quantiles_stein.round(2))
```

Bibliography

A. J. Baranchik. Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution. 1964. URL https://purl.stanford.edu/hq883mw7116.

W. James and Charles Stein. Estimation with Quadratic Loss. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 443–460. Springer, New York, NY, 1961. ISBN 978-1-4612-0919-5. doi: 10.1007/978-1-4612-0919-5_30. URL https://doi.org/10.1007/978-1-4612-0919-5_30.

Norman L. Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. Continuous univariate distributions. 2. Wiley, New York, 2. ed edition, 1995. ISBN 978-0-471-58494-0. Num Pages: 719.

Richard Samworth. Small Confidence Sets for the Mean of a Spherically Symmetric Distribution. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(3):343–361, June 2005. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2005.00505.x. URL https://doi.org/10.1111/j.1467-9868.2005.00505.x.

Mark J. Schervish. *Theory of Statistics*. Springer Science & Business Media, December 2012. ISBN 978-1-4612-4250-5. Google-Books-ID: s5LHBgAAQBAJ.

Charles Stein. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (1954/1955)*, 1954. URL https://digicoll.lib.berkeley.edu/record/112831. Num Pages: 197.